

[illegible]

QY 1763 aggcctcccaagatccatggaacacactactgaactctgcacactcccaagaaagaaac 1822
 Db 1548 AGGCTCCCAAGGCTCCATGGACACCTACTGAGACTCTGCCACTCCAGGAGAGGAAC 1697
 QY 1823 ctgaatcccccaaccctcccaactctggttgaagcaagagagagagagagagagag 1882
 Db 1698 CTAGCAATCCCAATCACTTCCACACTGCTGTTGAGCAAGACAGAGTGGGGAGCAACTGGT 1757
 QY 1883 ggtctgaactatctaggttccctccaggaagagagagagagagagagagagagagag 1942
 Db 1758 GGTCTGAGCTATCTGGGTCTCTTCAGAGGAGAGAGAGAGAG. CAGGAGAGCTCGAGGGT 1817
 QY 1943 ggccttccctgcttccagagacagagagagagagagagagagagagagagagag 2002
 Db 1818 GCTCTTCTGCTGCTTCTCAGACACAGAGGAGGCTCTGAGGTACAGGAGCTGAGAGGCTCC 1877
 QY 2003 tctgaagataattctgaagaaactaccccaagaaagagagagagagagagagagagag 2062
 Db 1878 TCTGAAGATAAATCTGGAGAGAACTGCTCCAGCAGGAGACCTCAGTTCAGAGGCTCAGCAGTG 1937
 QY 2063 ctgcacactgacag 2122
 Db 1948 CTGACCACTGACAGAGGCGCAGAGGTGGAGTGGCGCTGCTCCCGGATCAGGTAATCT 1997
 QY 2124 ggcagaggtcgaactgacctctctatctatctcttcttccctgagagctctgagctc 2182
 Db 1978 GCGCAAGGCTCAACTGGCGCTCTCTATCTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2057
 QY 2183 acctgacctgagctcttcaaccacacacacacacacacacacacacacacacacac 2242
 Db 2058 ACTGACCTGCTAGTCTCTTAAACCCACCACTATCCCAAACTCTCTCTCTCTCTCTCTCT 2117
 QY 2243 tctcttaccacacttacctatgagctctcaactctgagatctcaactctgagatctct 2302
 Db 2118 TCTCTTACCTACCTACTACTATPSSGCTCCAACTCCGATATCTCGCTCTCTCTCTCTCT 2177
 QY 2303 agctctccagctcttaccctgtatctccagccacacacacacacacacacacacacac 2362
 Db 2178 AGCTCTCCCGCTCTTACCTCTGTATCTCCAGCCGCTCCACTGACCATCTGTGACCTCTCC 2237
 QY 2363 ctgaccttgggacctccaccttggctccacacacacacacacacacacacacacacacac 2422
 Db 2238 CTGCCATGGGGCTCCACCTGCTGCTCAGCTTCAGAGCTTCAGAGGAGGAGCTCTCAGG 2297
 QY 2423 cttctccagaggtctcatcacctatctgagctctcaggtctcagctctctctctctctac 2482
 Db 2298 CTTCTCCAGGCTCTCATCACTATTCAGAGCTTCAGAGCTTCAGAGGAGGAGCTCTCAGG 2457
 QY 2483 tcccttaccagctgtgagctctcccttaccctgctctctctctctctctctctctctctac 2542
 Db 2358 TCCCTTCATTCGCGCTGTGCGCTGCTCTTACCTGCTCTTACCTGCTCTTCTCTCTCTCT 2417
 QY 2543 tgaagagctcttccacccctccacacacacacacacacacacacacacacacacacacac 2602
 Db 2418 TGGAGTCTCTCTCTCAACCTTCAGAGCTTCAGAGCTTCAGAGGAGGAGCTCTCAGG 2477
 QY 2603 ccttgaagcaactactctctgagagagagagagagagagagagagagagagagagagag 2662
 Db 2478 CTTTAAGCAACTACTCTGTGAAGTATTTTGTGACTGTTTCTGAGGAGGAGGAGGAGGAG 2537
 QY 2663 aataaatctctattaaac 2683
 Db 2538 AATAAATCTCTATTAAACCCG 2558

RESULTS 2

AX056675 AX056675 3476 bp DNA PAT 17-JAN-2001
 LOCUS
 DEFINITION Sequence 7 from Patent WO0075317.
 ACCESSION AX056675
 VERSION AX056675.1 GI:12309661
 KEYWORDS human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;

REFERENCE 1 (bases 1 to 3476)

AUTHORS Boistein, D.A., Goddard, A., Gurney, A.L., Smith, V., Walander, C.K. and

Wood, W.J

TITLE Composite maps and methods for the treatment of tumor

JOURNAL Patent: WO 0075317-A 7 14-DEC-2000;

FEATURES

GenBank, Inc. (US)

Location/Qualifiers

1..3476

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 679 a 1094 c 1092 g 611 t

ORIGIN

Query Match

Best Local Similarity 77.18; Score 2104.4; DB 6; Length 3476.

Matches 211; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 12 cgtcgtctctgcccct 21

Db 191 CGAATGTCCTCCAGGCT 250

QY 72 tgcagctgag 131

Db 251 TGCAGCTGCAAGCTGAGGCT 410

QY 132 gcaatgcccac 193

Db 311 GCAGTGTCTCCAGCCCAATCT 670

QY 152 cctctgag 251

Db 371 CCGTGTCT 430

QY 252 gagacagctcag 311

Db 431 GAGACAGCTCAG 490

QY 312 gctgagctcag 371

Db 491 CCGTGTCT 550

QY 372 gcccctcag 431

Db 551 GCGTGTCT 610

QY 432 aggcagagctgctgag 491

Db 611 AGGCAGAGTCTGCTGTGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 670

QY 492 gcaatgcccac 551

Db 671 GCGTGTCT 730

QY 552 tgcagctgag 611

Db 731 TGCAGCTGCAAGCTGAGGCT 790

QY 612 gcaatgcccac 671

Db 791 GCGAGCTCTCTGAGGCTCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 850

QY 672 gctatct 731

Db 851 GCTATGCT 910

QY 732 ccaacacag 791

Db 911 CCACCCCGAGGAGCT 970

QY 731 qccaccccgagacacgtctatgctgcttaccctgggggtatgagcaatgcatgctg 790
 Db 661 GCACCCCGGACGAGCTCTATGCGGGCTACCTTGGGGCTATGAGCAATGATGCTGCC 720
 QY 791 tggctgtggatcagacgtgaggtatccatccagacccacagagagcctgttacgga 850
 Db 721 TGCTGTGGATCAGACGCTGAGGTATCCCATCCAGACCCACAGAGGCGCTGTACGA 780
 QY 851 qacalggatggtctcccccgggtccggaactatggtgtggtggaccgagatgacttat 910
 Db 781 GACATGATGGCTTCCCGGGGTCCGGAACTATGCTGTGGTGGACCCGGATGACCTCTAT 840
 QY 911 qatqatgactgtatgctgaagacctaaatgagaaactgttccctgggtgacctccagag 970
 Db 841 GATGTCTACTGTATGCTGAAGACCTAAATGAGAACTGTTCCTGGGTGACCTCCAGAG 900
 QY 971 aagctgacatggaggaagcagcagcgtactgtccagagagaggggtgcagagattgcac 1030
 Db 901 AGCTGACATTTGGAGGAGTACGGGGGTACTGCCAGAGAGTGGGTGCAGAGATTGCCACC 960
 QY 1031 acggcccaactgtatgcagcctggtatggtgacctggacctgagccgagcagagggcta 1090
 Db 961 ACGGCCCAACTGTATGCAGCCTGGGATGCTGGCTTGGACCACTGCCAGCCAGGGTGGCTA 1020
 QY 1091 qrtqatgacagtgctgctaccccatcctcctacacccagcagcgtgtgtggggcttg 1150
 Db 1021 GCTGATGGCAGTGTGGCTACCCCATCTGACACCCAGCCAGCGCTGTGCTGGGGCTTG 1080
 QY 1151 cctgtgtcagagactctctctctctcccaacacagactggtctcccaatgacagagc 1210
 Db 1081 CTTGTGTCAAGACTCTCTCTCTCTTCCCTTCCCAACAGAGCTGGCTTCCCAATAGCACAGC 1140
 QY 1211 cgtctacacgtctactgtctcagagactggtcagcgtctctgacctcctgagcctcc 1270
 Db 1141 CCGTTCAAGCTCTACTGCTTCCGAGACTGGGCCAGCGCTTCTGCCATCCCTGAGCGCTCC 1200
 QY 1271 aacccagcctccacccacccctctatgagactagagctctcagcagcagagagacc 1330
 Db 1201 AACCCAGCGCTCCACCCAGCGCTCTGATGAGTACAGAGCTATCTGCACAGTGCACAGAGC 1260
 QY 1331 ctgagagaaactgagctgctcaggaacccagagagtgatccctggggccatctac 1390
 Db 1261 CTGAGAGAACTGAGCTGCTCTCAGGAAGCCACAGAGAGTGAATCCCGTGGGCCATCTAC 1320
 QY 1391 tccatcccatctgagcagcagcagctgagagctccactccagagagcctcagcagag 1450
 Db 1321 TCCATCCCATCTGAGGACCGAGAGGTTGGAGCTCCACTTCCAAAGACCCAGCGAGAG 1380
 QY 1451 gcccctaggcgtcctagaatttgaacacacatccatggtccgcccacaggggtctca 1510
 Db 1381 GCCCTAGGACGCTCTAGCAATTGAAACACAAATCCATGCTGACCCGCCACGGGGTCTCA 1440
 QY 1511 gaagagaaagttaagcattggagagagagagaatatagaagatgaagagagagagag 1570
 Db 1441 GAAGAGAAAGTTAAGCATTTGAGAGAGAGAGAGAAATATCAAGATGAAGAGAGAGAGAG 1500
 QY 1571 gaggagaaag 1630
 Db 1501 GAG 1560
 QY 1631 agagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1690
 Db 1561 ACCAGCCCGCGCTCAGCGCTCTCTCCCACTGAGCCAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1691 tccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1750
 Db 1621 TCCAGCGCTCAGAGAGGAG 1680
 QY 1751 tccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1810
 Db 1681 TCAGAGCTTCCAGCGCTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 QY 1811 agggag 1870

Db 1741 AGGAGAGAGAACTAGCAATGCCATACACCTTCACACTCTGTTGAGAGCAAGAGAG 1800
 QY 1871 gaggcaactggtggtcctgagctatcctggggtcctcagagagagagagagagagagag 1930
 Db 1801 GAGGCAACTGTTGCTCTGAGCTATCTGGGTCCTTCGAGGAGAGAGAGAGAGAGAGAG 1860
 QY 1931 agctcccaag 1990
 Db 1861 AGCTCCAGAGGTTGCTCTGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 1920
 QY 1991 ctgag 2050
 Db 1921 CTGAGAGGCGGCTCTGAGAGATAATCTGGAAGAACTGCCCCAGCAGGAGATCTAGTCCAG 1980
 QY 2051 gcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2110
 Db 1981 GCCCAGCCAGTCTGCCACCTGACAGAGCGTAGCCGAGGTGGAGTGGGCTGCTGCTGCTG 2040
 QY 2111 tcagqtaatttgcgcaagctc 2133
 Db 2041 TCAGCTGACATGTTGCTCCAGCGC 2063

RESULT 5
 AF228710 2878 bp mRNA 8K1 14-141-2000
 LOCUS Homo sapiens chondroitin sulfate proteoglycan BEHAB/brevican mRNA.
 DEFINITION soluble isoform, complete cds.
 ACCESSION AF228710
 VERSION AF228710.1 GI:10798900
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2878)
 AUTHORS Gary,S.C., Zerillo,C.A., Chiang,V.L., Gaw,J.U., Gray,G. and Hockfield,S.
 TITLE cDNA cloning, chromosomal localization, and expression analysis of human BEHAB/brevican, a brain specific proteoglycan regulated during cortical development and in glioma
 JOURNAL Gene 256 (1-2), 139-147 (2000)
 MEDLINE 20509959
 REFERENCE 2 (bases 1 to 2878)
 AUTHORS Gary,S.C., Zerillo,C.A., Gaw,J.U., Cheng-Moy,V. and Hockfield,S.
 TITLE Direct Submission
 JOURNAL Submitted (26-JAN-2000) Neurobiology, Yale University, 333 Cedar St., New Haven, CT 06520, USA
 FEATURES
 Location/Qualifiers
 1..2878
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q31"
 /tissue_type="normal cortex"
 48..2878
 /note="brain-specific; soluble isoform"
 /codon_start=1
 /product="chondroitin sulfate proteoglycan BEHAB/brevican"
 /protein_id="AAG2134.1"
 /db_xref="GI:10798901"
 /translation="MAQLFLPLIALVLAQAPAAALAUVLGDSSEDAFVRIRIAGDAP
 LQCVLGGLTIPCHVRLRPSPRAVLSRPRKWTFLSRGREAELVAVRGVVRVNE
 AYRFRVALPAYPSITDVLSALREIHPNDGIYRCEVHGIDUSSIAVEKVKGVFL
 YREGSARVAFSEGAQACARIGAHATPQLYAYLGVEQDQDAGWLSQIVRYPIQ
 TPREACYGDMGDFPVNRYVDFDIDVTCYADLNGELFLGDPPEKLTLEEARAY
 CUEGRAREATTGQLYAAWDGGLDHCSPGLADSVRYPIVTPSGGGGLPGVKTLFL
 FVNGTFPNKHSRENFYCPDQSAIPASNSPNSPAGSLAIVTVELELQ
 POPATSESGAIYSIPIMEDGGSGSTPDPAAPRTILLEFTQSMVPTGFSSEEG
 KALREERYEDEEEKREEREERERDEALWNPSELSPGPASLPTEPAQAQESLQ
 APARVLPQASPLDGESEASRPFRVHGPTETLPTPRENLASPSPTLYEAREVG
 EATGPELSGVPRGESEETGSGSEKAPSLPATRPEGTRELEAPSDNSGRTAPAGTS


```

Db 1998 GTCGCCAGGCC 2008      ||| |||
RESULT 6
LOCUS BTBREV 3259 bp mRNA MAM 24-MAY-1994
DEFINITION B.taurus Br. vican mRNA.
ACCESSION X75887
VERSION X75887.1 GI:452820
KEYWORDS brevican.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 3259)
Yanada,H., Watanabe,K., Shimonaka,M. and Yamaguchi,Y.
Molecular cloning of brevican, a novel brain proteoglycan of the
aggreacan/versican family
J. Biol. Chem. 269 (13), 10119-10126 (1994)
9419357
J-ARNAI.
MEDLINE
REFERENCE 2 (bases 1 to 3259)
AUTHORS Yamaguchi,Y.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1994) Y. Yamaguchi, La Jolla Cancer Research
Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
Location/Qualifiers
1..3259
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="brain"
/clone_lib="bovine brain lambda gt11 cDNA library"
112..2850
/gene="brevican"
112..2850
/codon_start=1
/gene="brevican"
/product="brevican"
/protein_id="CAA53481.1"
/db_xref="GI:452821"
/db_xref="SWISS-PROT:Q28062"
/translation="MAPLEFILLATVLAWIFVALALAEGLSSEDRARVRRTAGIAP
LQVLGGALVTPCHVHLRPSRRRAQSPRKWTFILSGRBAEVLVANGLRVKVSE
AYRFRVALPQVSLDVSILVSELPNDSGYRCEVCHGIDDSDAVENKVGWVFL
YRESARYAFSPAGAGAACARIGARLATPEOLYAAVLGGVEQDCADWLSDQVRYPIQ
TPREACVGMDFGPVRYNGWDFDLDLYCYAEELNGELF LGAPDKLILEARTY
CQGRGAKIATTCGLYAADGGLDKSSGSLSDGSVRYPIVTPSGGGGIPGVKTLFL
FPQGTGPNKHSRFNVYCFRDSAPSAIPEAANPASHLASUALKAIIVTETLEELKL
PQAVESESGAIVSIPIIEUGGGSTPEDEAPRPTLLFEFQSVTPPLGSSSEEG
KVLQOEKRYGEEKEEREVEDEALWAPSLSSLDPEAPLTPFPVPESELTOA
SPVRAALQPGVSPDPYDEAPRPRVLGPPTKLTPTREGNLASPPSTLVGAREI
BEETGPELSPGAPGSEETGSESDAPSLPATRAPGTRELETPSEENRRTVPGAT
SVRAQVLPDTSASRGVAVAPSGDCVSPCHNGGTCLREEGVCRLCLLPYGGNHC
DLGHRCSPGMDATOGACYKHFSAKRSWBAENKCRMYGAHLASISTPEQDFINRY
REQTWILGNDRITIEDFLWSDGVPILYENNPQCPUSYFLSGENCYVMVWHQGWMSD
VPCNTGLSTKMTKXILVSGPPPELPLAEVFGPRPLRYEVDIVLYRCREILTORNPL
IRQENGRWGLPQISCVPRRPRARALRPEVQSEGRWRLVCHWKARLNPSPAPGP"
BASE COUNT 624 a 1026 c 991 g 618 t
ORIGIN
Query Match 58.0%; Score 1582.4; DB 4; Length 3259;
Best Local Similarity 86.9%; Pred. No. 1.5e-300;
Matches 1772; Conservative 0; Mismatches 246; Indels 22; Gaps 2;
QY 116 gaagcgcacaaacttcagtcgcgcgacccacacccacgcgcctggat----- 162
Db 36 GGAGGGGCAAACTTCGCGTGCCTCAACCCCGCGTCTCCTCACTTGCCTGGCGTCCC 95
QY 163 -----agcctgcagcagtcgcgcagctattctacccctacttcagcaccctcttgc 216
Db 96 TCCCTCATCTCTGACAGATGGCCCGCCACTGTCTCTGCCCTCTCTGCGCAACCCCTGCTTGC 155

```



```

QY 2423 cctctcaccagctctctcat--actatttgcagccltcagagctcaggctatttttctact 2480
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2271 TCTCTTCTGTCGTCCTTCACAGCCCTTCAGACCTTCAGACCTTCAGACCTTCAGACCTTC 2330
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2481 actcctctcaccagctctctcat--actatttgcagccltcagagctcaggctatttttctact 2540
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2331 TCCCTCTCCTTCGTCCTTCCTTCATCCTTCATCCTTCATCCTTCATCCTTCATCCTTCATC 2390
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2541 cctggagctcctctcaccagctctcagcagctcagcagcagcagcagcagcagcagcagcag 2600
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2391 CCTGAGACACCATCCTGTCCTTC-ACCTCTGGGACACAGAAATA---GTACCCCAAG 2445
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2601 accttaagacactctctgagctatttttgcagctatttttgcagctatttttgcagctatttt 2660
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2446 AGCCTTAAGACACTTACTTCTGTCAGATATTTTCTACTGTGTTTATGGAACACAGCCCTTG 2505
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2661 gaataaactctatttaaacgcctttgta 2689
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2506 GAATAAATCTGATTAAACTCTTTGTA 2534
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AL590666
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-66D17, *** SEQUENCING IN
PROGRESS ***, 11 unordered pieces.
ACCESSION AL590666
VERSION AL590666.6 GI:13992142
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
McLay, K.
Direct Submission
Submitted (03-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerquest@sanger.ac.uk
On May 8, 2001 this sequence version replaced gi:13990633.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA66D17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big dye; 100% of reads
Consensus quality: 158757 bases at least Q40
Consensus quality: 159368 bases at least Q30
Consensus quality: 159820 bases at least Q20
Insert size: 160318; sum-of-ctigs
Insert size: 149188; 13.8% error; agarose-fp
Quality coverage: 9.28x in Q20 bases; sum-of-ctigs Quality
coverage: 9.97x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 25475: contig of 25475 bp in length
* 25476 25575: gap of 100 bp
* 25576 41767: contig of 16192 bp in length
* 41768 41867: gap of 100 bp
* 41868 44284: contig of 2417 bp in length
* 44285 44384: gap of 100 bp

```

```

* 44385 54678: contig of 10294 bp in length
* 54679 54775: gap of 100 bp
* 54779 63396: contig of 8518 bp in length
* 63397 77898: contig of 14502 bp in length
* 77899 77998: gap of 100 bp
* 77999 81999: contig of 4001 bp in length
* 82000 82099: gap of 100 bp
* 82100 91953: contig of 11854 bp in length
* 91954 94053: gap of 100 bp
* 94054 96641: contig of 2578 bp in length
* 96632 96731: gap of 100 bp
* 96732 101107: contig of 4376 bp in length
* 101108 101207: gap of 100 bp
* 101208 161318: contig of 60111 bp in length.
FEATURES
    source
        1..161318
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-66D17"
            /clone_lib="KPC1-11.1"
            1..25475
                /note="assembly_fragment:04075"
                fragment_chain:1
                clone_end:SP6
                vector_side:left
                25576..41767
                    /note="assembly_fragment:02053"
                    fragment_chain:1
                    41868..44284
                        /note="assembly_fragment:02777"
                        fragment_chain:1
                        44385..54678
                            /note="assembly_fragment:04258"
                            fragment_chain:1
                            54779..63396
                                /note="assembly_fragment:01042"
                                fragment_chain:1
                                63397..77898
                                    /note="assembly_fragment:03185"
                                    fragment_chain:1
                                    77999..81999
                                        /note="assembly_fragment:03976"
                                        fragment_chain:1
                                        82100..93953
                                            /note="assembly_fragment:00892"
                                            fragment_chain:1
                                            94054..96641
                                                /note="assembly_fragment:02368"
                                                fragment_chain:1
                                                96732..101107
                                                    /note="assembly_fragment:02159"
                                                    fragment_chain:1
                                                    101208..161318
                                                        /note="assembly_fragment:01356"
                                                        fragment_chain:1
BASE COUNT 37683 a 40929 c 41273 g 40430 t 1003 others
ORIGIN

```

```

Query Match 45.0%; Score 1228.4; DB 2; Length 161318;
Best local Similarity 99.9%; Pred. No. 2.5e-231;
Matches 1229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1465 cctagaatttgaacaacatccatggtaccgccacgggqtlctcagaagcaggaataa 1524
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 CCTAGAATTTGAACAACATCCATGGTACCGCCACGGGTTCTCAGAGAGGAGGTAA 921
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1525 ggcatttgaaggaagaagaataatgaagatgaagaagaaggagggaagaaca 1584
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 GGCATT:3GAGGAAGAAGAATAATGAGATGAAGAAGA:..AGGAGGAAGAAGAACA 981
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
QY 1585 ggaagaggtgagatgaggtctgtggacatggccnagagaggtcagcaagccgagccc 1644
Db 982 GGAGGAGGTGGAGATGAUGCTCTGTGGCATGGCCAGCGAGTTCAGGAGCCGCGGCC 1041
QY 1645 tgaagccctctccrctcagccagccagccagccagccagccagccagccagccagcc 1704
Db 1042 TGAGGCGCTCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1101
QY 1705 aaggcagctcctcagccagccagccagccagccagccagccagccagccagcc 1764
Db 1102 AAGGCGAGTCTGCGAGCTGGTGCATCAGCACTCTCCAGCTCCAGGAGGAGGAGGAG 1161
QY 1765 gctcccaaggtccatggacacactactgagactctgcccactcccccagcagagaaac 1824
Db 1162 GCCTCCAGAGGTCCATGGACACCTACTCAGACTCTGCCACTCCAGGAGGAGGAGGAG 1884
QY 1825 agcatcccccacactccactctgttgaagcagagagagagagagagagagagagag 1884
Db 1222 ACCATCCCATCAGCTTCCACTCTGGTTAGGCGAGAGAGGAGGAGGAGGAGGAGGAG 1281
QY 1885 tctaaagctatctgggtccctcagagagagagagagagagagagagagagagagagag 1944
Db 1282 TCGTGAATATCTGGGTCCCTCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1341
QY 1945 ccttccctgtctccagccacacagggcccttgaaggtaccagagagagagagagagag 2004
Db 1342 CCGCTTCCCTGCTTCCAGCCACAGCGGCGCCCTCAGGGTACCAGGAGGAGGAGGAGGAGG 1401
QY 2005 tgaacataatcttgaagagactgccccagcagcagcagcagcagcagcagcagcagc 2064
Db 1402 TGAAGTAATTCGGAAGAGAACTGCCGCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 1461
QY 2065 ggcactgaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2124
Db 1462 GCGACATGAGAGCGCCAGCGAGGTGGAGTGGCGCTGGTCCCGCATCAGGTAAATCTGCG 1521
QY 2125 caaggtctaaactgacctctctactactctctctctctctctctctctctctctctct 2184
Db 1522 CCAAGGTCTAACTGCGCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1581
QY 2185 ctgaccltgaagccttaccacacacacacacacacacacacacacacacacacacac 2244
Db 1582 CTGACCTCTACTCTTTACCCACACACACACACACACACACACACACACACACACAC 1641
QY 2245 tcttaccacacacacacacacacacacacacacacacacacacacacacacacacac 2304
Db 1642 TCTTACCACCTCTACTCTATGGGTCTCCATCTCCGATATCCACCTTGTGGGTATCTCAG 1701
QY 2305 ctctcccgctttacccttgatccnagcccgccagctgaacatctctgaccccttccct 2364
Db 1702 CTCCTCCGCTCTTTACCTCTGTATCCAGCCCGCCGACCTGACCATCTGTGACCCCTCC 1761
QY 2365 gccattgggcccctccactctggtccacatctgcccagcccccagcagcagcagcagc 2424
Db 1762 GCGATTGGGCGCCCTCCACCTCTGGGCTGACATCTGCCAGGCCACACAGCATCTCTCAGGCC 1821
QY 2425 tctccagggctctctacacactatttgcagcttcagggctcagggctcagggctcagggct 2484
Db 1822 TCTCCAGGCTCTCATCACTCTATTCAGGCTTTCAGGGCTCGGGCTATTTTCTCACTACTC 1881
QY 2485 ccttcatagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2544
Db 1882 CCTTCATCTCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1941
QY 2545 ggaatcccttccacccctcaactccagagctccagagagagagagagagagagagagag 2604
Db 1942 GGAGTCCCTCTCACCCCTCAACCTCCGGAGTCCAGGAGAACCGTACCCGCCACAGAGCC 2001
QY 2605 ttaagcaactactctgttgaagatatttttgaactgttttgaagaaacaaacacacacac 2664
Db 2002 TTAAGCAACTACTCTGTGAAGTATTTTGAAGTATTTTGAAGTATTTTGAAGTATTTTGA 2061
QY 2665 taaatctctattaaacccgcttttgaacaa 2694
```

Db 2062 TAAATCTCTATTAAACCTTTTGAACAA 2091

RESULT 9
AL365181/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL365181 178753 bp DNA HTG 17-MAR-2001
Homo sapiens chromosome 1 clone RP11-284F21, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
AL365181
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
J (Buses 1 to 178753)
McLay, K.
Direct Submission
Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13473967.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA284x21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator ET-amersham; 13% of reads
Dye-terminator Big Dye; 86% of reads
Consensus quality: 174554 bases at least Q40
Consensus quality: 175989 bases at least Q30
Consensus quality: 176840 bases at least Q20
Insert size: 177453; sum-of-ctctigs
Insert size: 189736; 1.5% error; agarose-1p
Quality coverage: 6.8x in Q20 bases; sum-of-ctctigs quality
coverage: 6.67x in Q20 bases; agarose-1p

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 46845: contig of 46845 bp in length
46846 46945: gap of 100 bp
46946 52507: contig of 5562 bp in length
52508 52607: gap of 100 bp
52608 58440: contig of 5833 bp in length
58441 58540: gap of 100 bp
58541 61563: contig of 3023 bp in length
61564 61663: gap of 100 bp
61664 66601: contig of 4938 bp in length
66602 66701: gap of 100 bp
66702 100522: contig of 33821 bp in length
100523 100622: gap of 100 bp
100623 111839: contig of 11217 bp in length
111840 111939: gap of 100 bp
111940 129656: contig of 17717 bp in length
129657 129756: gap of 100 bp
129757 140052: contig of 10296 bp in length
140053 140152: gap of 100 bp
140153 143799: contig of 3647 bp in length
143800 143899: gap of 100 bp
143900 155397: contig of 11498 bp in length
155398 155497: gap of 100 bp

* 155498 162028: contig of 6531 bp in length
 * 162029 162128: gap of 100 bp
 * 162129 166344: contig of 4216 bp in length
 * 166345 166444: gap of 100 bp
 * 166445 178753: contig of 12309 bp in length.

FEATURES

source

1..178753
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-284F21"
 /clone_lib="KPCI-11.1"

misc_feature

1..46845
 /note="assembly_fragment:00916
 fragment_chain:1
 clone_end:SP6
 vector_side:left"

misc_feature

46946..52507
 /note="assembly_fragment:01834
 fragment_chain:1"

misc_feature

52608..58440
 /note="assembly_fragment:02483
 fragment_chain:1"

misc_feature

58541..61523
 /note="assembly_fragment:02948
 fragment_chain:2"

misc_feature

61664..66601
 /note="assembly_fragment:01794
 fragment_chain:2"

misc_feature

66702..100522
 /note="assembly_fragment:01092
 fragment_chain:2"

misc_feature

100623..111839
 /note="assembly_fragment:01799
 fragment_chain:2"

misc_feature

111940..129656
 /note="assembly_fragment:01722
 fragment_chain:3"

misc_feature

129757..140052
 /note="assembly_fragment:01996
 fragment_chain:3"

misc_feature

140153..143799
 /note="assembly_fragment:01161
 fragment_chain:3"

misc_feature

143900..155397
 /note="assembly_fragment:01410"
 fragment_chain:4"

misc_feature

155498..162028
 /note="assembly_fragment:00273
 fragment_chain:4"

misc_feature

162129..166344
 /note="assembly_fragment:02309
 fragment_chain:4"

misc_feature

166445..178753
 /note="assembly_fragment:02469
 fragment_chain:4
 clone_end:T7
 vector_side:right"

BASE COUNT 43774 a 45152 c 45060 g 43462 t 1305 others

ORIGIN

Query Match 45.0%; Score 1228.4; DB 2; Length 178753;
 Best Local Similarity 99.9%; Pred. No. 2.5e-231;
 Matches 1229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1465 cctagatttgaacacaaatcctatggtaccgcccggtttctcagaagagaaggttaa 1524

Db 21717 CCTAGAAATTGTAACACAAATCCATGTTACCGCCCGCGGGTCTCCAGAGAGGAAGTAA 21658

QY 1525 ggcattggaggaagaagaataatgaagataaagaagaagaagaagaagaagaaga 1584

Db 21657 GGCATTGGAGGAGAGAGAGAAATATGAGATCAAGAGAGAGAGAGAGAGAGAG 21598

QY 1585 ggaagaatggagaatgagactctgtgggcattggcccagcagagctcagcagagccagagcc 1644

Db 21597 GGAGGAGGTGGAGGATGAGGCTCTGTGGGATGGCCAGGAGGCTCAGCAGTCTGGAGCC 21538

QY 1645 taaagactctctcccactgagcccaacacccacagagaagtcaactctcccacacccacag 1704

Db 21537 TGAGGCGCTCTCTCCCTCACTGAGCCACAGCCACAGGAGAGAGTACCTCCACAGGAGTACG 21478

QY 1705 aagagcagtcctcagcagacttgatgatacaccacttctcagatggagactcagaagacttcag 1764

Db 21477 AAGGCGAGTCTGCAGGCTGGTGCATCACCATTCTCTGATGGAGATCAGAGATCTCAG 21418

QY 1765 gacctcraagggtccatgagaccactactgaactctgcccactcccccagagagagagag 1824

Db 21417 GGTCTCAAGGGTGCATGGAGCTAGCTGAGACTGTGCCAATCCCCAGAGAGAGAAATCT 21358

QY 1825 aacatcccacacacttccactctggttcagagcagagagatggagagagagagagag 1884

Db 21357 AGCATCCGCAACACTTCCACTCTGGTTGAGGCAAGAGAGGTGGGGAGAGGCAAC 21298

QY 1885 tctcagactatctagggtccctcagagagagagagagagagagagagagagagagagag 1944

Db 21297 TCTCAGCTATCTCGGGTCCCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21238

QY 1945 cccctccctgcttccagccacagagccctcagagatccagagagagagagagagagagag 2004

Db 21237 CCGTTCCCTGCTTCCAGCCACAGGGCCGCTGAGGGTACCGAGAGAGTGGAGAGAGAGAG 21178

QY 2005 tgaagataatctcgaagaactgcccagagagagagagagagagagagagagagagagag 2064

Db 21177 TCAAGATATCTCTGAGAGAGAGTCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21118

QY 2065 gcccacccacagagccag 2124

Db 21117 GPCCATGACAGCGCCAGCGAGGTGAGTGGCGTGGTCCCGCCATCAGTAATCTCAG 21058

QY 2125 caaaggctcaactgacctctctactactactcttcttctcccccagagctctgagatcac 2184

Db 21057 CCAAGGCTCAACTGGCTCTCTATCTACTCTTTCTTCCGCTCTGACATCTGGGTAC 20998

QY 2185 ctgacctatagacttlaacccacacatcccaaatctctccctctcttggctctctc 2244

Db 20997 CTGACCTGTAGTCTCTTAAACCCAGCATTAACCAAAACCTGCTTGGCTTTGGCTTCAATC 20938

QY 2245 tcttccacacctctacctatgggtctctccatctccagatataccaccttgtagatctctcag 2404

Db 20937 TCTTACCCACCTCTACCTATGGTCTCCAAATCTGAGATATCCACCTTGTTGGGTATCTCAG 20878

QY 2305 ctctccgctctcttaccctatgactccaccccacacacttaacactctgtagacctctccct 2364

Db 20877 CTCTCTCTGCTTTTACCTCTGATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 20818

QY 2365 ggcattggagccctccacctatgctccatctctccagagcccccacagagatctccagagcc 2424

Db 20817 GGCATTGGGCGCTCCACCTCTGCTCAATCTCTCCAGCCAGCCACAGAGATCTCTCAGAGCC 20758

QY 2425 tctcraagggtccatccatctcagagagagagagagagagagagagagagagagagag 2484

Db 20757 TCTCCAGGGTCCCTCATCACTACCTATTCCAGGCTTCCAGGCTCGGCGCTATTCTTCACACTC 20698

QY 2485 cttctacccgctctgctgctgctctctttagctgctctctctctctcagagagagagagag 2544

Db 20697 CTTCTATCCGCTCTGTGGCGGTTCCTTTAGTGGCTCTCTATTGATCTCTCAGGAGAGGCTG 20638

QY 2545 ggaactccctctccagctccacactccagagatccagagagagagagagagagagagagag 2604

Db 20637 GGAGTCCCTCTCTCAGCCCTCAACCTCCGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 20578

QY 2605 tlaagcactactctctgagaagtattttttagctggtttctcatggaagaagaagagagag 2664

Db 20577 TTAAGCAACTACTTCTCTGTAATTAFTTTTGTACTGTTTCTATGAAATCAAGAGAGAGAG 20518

QY 2665 taaatctctattaaacgcttttgaaccaa 2694

QY 629 aaggt caaagggatctcttctctaccagagagctctaccagctatgcttctctctt 688
Db 513 AAGGTCAAGGGGTGCTCTCTCTACCGAGAGGGCTCTCCCGCTATGCTTCTCTTC 572
QY 689 tctgggcccagagagctctgtcccgcatctgagcccaatgcacacccagccagagagc 748
Db 573 GCTGGAGCCAGGAAGCTGTGCTCCGATCG-GAGCCGGAATTGCCACCCCTGACGAGTG 632
QY 749 tatccacccctacncttgggagctatga :aatygaatctgggtggtctgctcagatcagacc 808
Db 633 TATGCTGCCCTACCTCGCGGCTATGAACA :tgtgatcctggctggctgtctccaccacac 692
QY 809 gtgaggtatcccatccacacccacagag :acctgttaacagacacatgagatgctctccc 868
Db 693 GTGAGGTACCCCAATCCAGAACCCACGAGAGCCCTGTATGGAGACATGATGGCTTACCT 752
QY 869 gggatcccaactatggtgtgagcccgatgaacctctatgatgtctactgtatgct 928
Db 753 GCAATGCGGAATTACCGAGTGGTGGCTCTCATGATCTACGATGCTTACTGTATGCC 812
QY 929 aagaactaaatggaactgttcclygtgaacctccagadaaactgacattgagaa 988
Db 813 CAAGACCTAAATGGAGAAGTGTCTTAGTGGCTCCCGGCAAGCTGAOCTGGAGAG 872
QY 989 acacgagctactgcagagaggggtgcagagatgccacacagggcccaactatgca 1048
Db 873 GCTCGGACTACTGTCTGGAAGCGGCTGCTCAGATGGCTAGCAGCGGCACTATACGG 932
QY 1049 qcttggatggtgagcactgcagccagagtgatgactgagtgagcagctgagc 1108
Db 943 GATGGAATGGCGGCTTGACAGATGTAGGCTTGGCTGGCTGATGGCAAGTGTGCG 992
QY 1109 taccatcagccacagcagctctgctgagggcttgcctggttgcctggttgcagactctc 1168
Db 993 TACCCATATACACCCAGCAACGCTGTGGGGAGGCTGCCAGAGCTCAAGACCCCTC 1052
QY 1169 tctctcttcacacacactgcttcccaatgaacacagctctcaactctactgc 1228
Db 1053 TTCTCTTTTCCCAACAGACTGGCTTCCCGAGCAAGCAGTTCATGCTACTGC 1112
QY 1229 tccagagctcggccagcctctgcatcctcctgaggtcccaacacagcctccaaacca 1288
Db 1113 TTCCGAGACTCTGCCATCCCTCTGCTCTCTGAG- - - - -GCTCCAGCCCA 1160
QY 1289 qctctgattgactagaggtctatcgtacagatgaacagacccctgaggaactcagctg 1348
Db 1161 GCCTCTGATGGACTAGAGGCCATTGTCACAGTCACAGAGAACTGGAGGAATCCAGTGG 1220
QY 1349 cctcagagaccacagagagtgaaatccgtggggctatctactcctcctcctcagag 1408
Db 1221 CCTCAGGAAGCTGTGAGAGCGAGTCTGCTGGCGGATCTACTCCATCCCATCAGAGA 1280
QY 1409 gacgagggggtggaactcctcctcagaaacccacagacagagccctagagcctccta 1468
Db 1281 CATGGGGAGAGGAGGCTCTACCCCAAGAACCCAGAGAGGCCCCCAGAGCTCCTCTTA 1340
QY 1469 gaattgaacacacatccatggtacccgcccaggggttctcagagaggaagtaaggca 1528
Db 1341 GAATCAGAAACCAATCGTTGCACCACTTACCAGCTCTCAGAGAGGAGCGGAGGCT 1400
QY 1529 ttgaagaaagaagaatatgaagatgaagaagaagagggaggaagaagagagag 1588
Db 1401 CTGGAGGAAGAAAGAT- - - - -TCMAAGACACAGAGACTCCGAGGGA 1445
QY 1589 gaggtgagagatgagctctctgtggcatggccacagcagctcagaccccgccctgag 1648
Db 1446 GAGAAGGACAGGAGACCTGTGGGTGTGGCCACGAGGCTCAGAGCCCTCTCCCTACT 1505
QY 1649 gctctctcccactgagccagcagcccagagagagagtagctctcctcagggcagcagag 1708
Db 1506 GGCTTGGAAACAGAGCACTCACTCTCCAGGTGTCCCA- - - - -CGAGCCAG 1553

QY 1709 ggaagttcagagctggtgcatacaccacttctctgattgagagattcagaagcttccagact 1768
Db 1554 GGAATTCTACAGCTGGGTGCATCACC- - - - -TCTCTCAGAGGCT 1592
QY 1769 ccaaggggtccatgaacacactatcgaactctgcccactcccaagagagagagactatca 1828
Db 1593 CCAAGGGGTCCATGGACCGGCTGCAGAGACTTTTGCAACCCCTCAAGGAGGAGGCTTCA 1652
QY 1829 tcccacatcactctcactctggttaagcacaagagaggtggggagagagcaactgagatct 1894
Db 1653 TCT- - - - -ACTCCAGATGGGCAAGAGAAGTAGCGGGGAAATTCGAGAGCT 1760
QY 1889 gagttatctggggtctctctgagagagagagagagagagagagagagagagagagagag 1942
Db 1701 GAGCTCTCTGGGTTCCTC- - -GAGAAACGAGAGAGGAGGAGAGCTCTCAGCTTGAGAGAT 1757
QY 1943 gcccctctctctgctctcagacacacagggccctgaggtaccagagagagagagagagag 2002
Db 1758 GCGCTTCCCTCTCCAGCGACATGGCCCTCTGTGGTACCAGGAGCTCCAGAGCTCC 1817
QY 2003 tctaaagaatctctggaagagactggtcagcagcagagagagagagagagagagagag 2052
Db 1818 TCAGAAGAGAGAGTTCGAGAGACTCTTCTGACAGGACATCATGTCAGAGGCTCAGGCACTG 1877
QY 2063 ctgcccactgacagcgcacgtgaggtgagatgagcctgaggtaccagagagagagagag 2122
Db 1878 CTGCCCACCGAGAGTGCAGCTGAGTGGAGTGGCTGTGGTCTGCTCATCAGAGTGAATGT 1947
QY 2123 gcccagagctc 2133
Db 1938 ATCCCAACCC 1948
RESULT 12
RNALPBR-
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
CDS
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="lambda gt10, lambda gt11"
/tissue_type="brain"
/codon_start=1
/product="aggregan like protein/ u-revican"
/protein_id="CAA56255.1"
/db_xref="GI:509397"
/db_xref="SWISS-PROT:P55068"
/translation="MPLLSLIIAALVLTQAPAAADLDKEDSSDRAFRVRIQAAGL

KGVLDGALALPCHVHHILRPPIRSKRAAGPFRVKWIFLISGDSREVEVLVARGLRVKVNEA
 YRFRVALPAYPASLTLSVLSLRLPNDSVVRCEVQHCIDISSDAVENVRKGVVFLY
 REGSARYAFSEAGAEACARIGARIATPEQLVAAVLGGYECODAGWLSLQUTRYPLON
 PREACYGDMXYPGVRNVTGVVDDLYCYCAEDLNGELFLGNPCKLTWEERADYC
 LERGAUJASTGOLYANNGLDNCSPWLDGSRVXPIIPIPSQRCGGGLPGVKTLFLF
 PRGTGPPSKNRHNVYCFRDSAHPSAFSEASPDSGLEAIVVTEKLEIQLJPOEAV
 ESRSGAIIPTIEDGGSSSTPEDAPAKTPLESTQVAPPTGCSBEEGEALAE
 ERFKDTPTKEKEQENIMWNPTLSPLPTGLESHLSQVSPPAQAVLQVGASPS
 PPPRVHGTVTETLQPPGEGSLSTPDGAREVGGTSPGSPVPREREAGSSLED
 GSPLLPETWAGTREVETSEKSGRTVLGTISVQAQVPLPTUSASRGGVAVAPSSG
 DGILSPCHNGTCLLEKEGFRCLCVPGXGDLCDWGLHFCSPWEFPQACACHFSR
 RSKAEASORCALCAHLTISCTPEBDQFVNDRYREYOMGLNDRTIEGDFKMSDPPPL
 LYENNNPQSPSYFLSGENCVWVHDDQOWSDPCNTHLSYTKMGLVSCGPPQLP
 LAGTIFDRPLRYAVDTVLYRCRDGLAQRNPLIRCOENGLWEAPQISVQVPRPARAL
 RNTAPEGRGQLPRQKALLTPSSL"

mat_peptide 173..2758
 misc_feature 208..579 /product="aggrecan like protein/ brevicin"
 repeat_region 586..882 /note="1q fold"
 repeat_region 883..1186 /note="proteoglycan tandem repeat 1"
 repeat_region 1187..1207 /rpt_type=TANDEM
 repeat_region 1208..1209 /note="proteoglycan tandem repeat 2"
 repeat_region 1210..1211 /rpt_type=TANDEM
 repeat_region 1212..1213 /note="BGF repeat"
 repeat_region 1214..1215 /note="BGF repeat"

Query Match 42.5%; Score 1160.4; DB 10; Length 2868;
 Best Local Similarity 77.2%; Pred. No. 9.8e-218;
 Matches 1534; Conservative 0; Mismatches 376; Indels 78; Gaps 7;

QY 149 cccacacccgtggtgagctgagcagatgcccacagctattcctgcccctgctgagacccgtg 208
 Db 86 CACAAAGTTCCTCAAGTTTGGCAGCATGCCATTCCTCTGTCCTGCTGCCAGCTTC 145

QY 209 qtcctgacccagctccgacgttlaqacagatattctgaagagagacagctcagaagac 268
 Db 146 GTCTGTAGCCAGACCCCTGCAGCCCTGCTGATACCTGAAAGAGACAGCTCAGAGGAT 205

QY 269 cgcacttttcgtgctgacgtcagcggcagcagcagcagcagcagcagcagcagcagc 328
 Db 206 CGAGCCCTTGGGTGGCAGTC---GGTGGCGCCAGCTGAGGCGGTGCTGGCGGTGCC 262

QY 329 cttacacatccctgacgtccactacgtcagcagcagcagcagcagcagcagcagcagc 388
 Db 263 CTGGCCATCCCATGCGCAGCTCCACACCTGAGGCGCGCGCCAGCGCGCGCGCGCGCG 322

QY 389 gqctctcgcaggtcaagtgacttctctgctccagagcagcagcagcagcagcagcagcag 448
 Db 323 GGCTTTCGCCGAGTCAATGAGCCCTTCTCTGCGGGGACCGGAGGTGGAGGTGCTGGTG 382

QY 449 gdcgagaggtgcgctgcaagtgtaacagagcctacaggttccgcggtgagcagcagcagc 508
 Db 383 CGCGCGGGCTCGCGTCAAGTAAACGAGGCTATCGGTTCGCGGTGGCGCTGCCCTGCC 442

QY 509 taccagagctgctcaccagagctccctgagcagcagcagcagcagcagcagcagcagcagc 568
 Db 443 TACCCGCCATCGCTCAGAGATGTCTTTAGTATTGAGCGAAGCTCGCGGCCCATGATGCC 502

QY 569 gqatctatcctgagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 628
 Db 503 GGGGTCTATCGCTGGAGGTCCAGACGGTATCGACGACAGCAGTGTGCTGGAGGTC 562

QY 629 aaggtcaagaggtgctgttctctaccagagagcagcagcagcagcagcagcagcagcagc 688
 Db 563 AAGGTCAAGAGGTGCTTCTCTCTACCGAGAGGGCTCTGCCCTATGCTTCTCTCTTC 622

QY 689 tctgagagcagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 748
 Db 1643 CCAAGGCTCCATGGACCGACGTGAGAGACTTTCCAAACCCCGGAGGAGGAACCTTCACA 1702

Db 623 GCTGGAGACCCAGGAAGCCCTGTGCTCCCATCGGAGACCCGAAITCCACACCTCTGAGAGAGTG 682
 QY 749 tatqgccttaccttgggggctatgagcaatattgatctqctggtgtgttgcagatcagacc 808
 Db 683 TATGCTGCTACCTCGCGGCTTATGAACASTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 742
 QY 809 gtdaggtatcccatccagacccacagagagcgttgttaccagagacatcagatcagcttccc 868
 Db 743 GTGAGGTACCCCATCCAGAACCCACAGAAAGCCTTTATGAGACATGAGAGTAACTCT 802
 QY 869 ggggtccggaactatggatgggtggagcccgagataacotctatgtagtactgttactct 928
 Db 803 GGAGTCCGGAATTAACGAGTGGTGGTCTGTGATGATCTACGATGCTACTGCTTATGCT 862
 QY 929 gaagacctaaatggagaaactgttcttggatgacccctccadadaaactgacatctaaagaa 988
 Db 863 SAAGACCTAAATGAGAACTGTTCCTAGGTGACCTCCGCGCAAGCTGAGTGAAGTGAAGAG 922
 QY 989 gcaagggcgtactcgcagagagaggggtgcaagagattgcacacacagagagacattatgca 1048
 Db 923 GCTCGGGACTACTCTCTGGAACCGGTGCTGATGCTGCTGATGCTGCTGATGCTGCTGCT 982
 QY 1049 gcttgagatgggtgacccactgcagccacagagatgggttaactaatgacatgacatgac 1108
 Db 983 GCATGGAATGCGCGCTTGCACAGATCTAGCCCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1042
 QY 1109 taccacatgltacacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1168
 Db 1043 TACCCCATATCATACGCGCCAGCCAGACCTGTGGGAGAGGCTGCCAGAGATCAAGAGCTTC 1102
 QY 1169 ttccttctcccaacagactggcttcccataatgacacacagcagcagcagcagcagcagcag 1228
 Db 1103 TTCTCTTTTCCCAACAGACTGGCTTCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1162
 QY 1229 ttcacacactcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1288
 Db 1163 TTCCCAAGACTCTGCCATCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1210
 QY 1289 gctctatgactagaactatgltacagatqaragagactctgagagaaactgactatg 1348
 Db 1211 GCTCTGATGGACTAGAGAGCAATTCACAGTGAACAGAGCTGGAGAGTGGAGAACTGAGTTG 1270
 QY 1349 cctcagaagacacag 1408
 Db 1271 CCTCAGAGAGCTGTGAGAGAGAGAGTCTGCTGGGCGGATCTACTCTATCCCATCACAGAA 1330
 QY 1409 gacgagagaggtggaactcactccagaaacacacagagagagagagagagagagagagag 1468
 Db 1331 CATGGGAGAGAGAGAGAGTCTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1390
 QY 1469 gaatttgaacacaaactcctatggtaccgacacagaggttctcagagagagagagagagag 1528
 Db 1391 GAATCAGAAACCCATCTGCTGACACAGCTACCGGGTCTCTCAGAGAGAGAGAGAGAGAG 1450
 QY 1529 ttgag 1588
 Db 1451 CTGAGAGAGAGAGAGAGAT-----TCAAGACACAGAGACTCCGAGAGGAA 1495
 QY 1589 gaggttgaagagatgaagcttctatggagatggccagagagcagcagcagcagcagcag 1648
 Db 1496 GAGAAGGAGAGAGAGAGAGTCTGCGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGT 1555
 QY 1649 gctctctcccactgacccagagcagcagagagagagagagagagagagagagagagag 1708
 Db 1556 GGGTTGGAAGACAGAGTCTACTCTCCAGAGGTGTCTCCCA-----CCAGCCCAAG 1603
 QY 1709 gcagctcctgagccttgggtcactcaccacttctctgagagagagcagagagcagcagcag 1768
 Db 1604 GCAGTTCTACAGTGGGTGATCACC-----TTCTCCCAAGGCT 1642
 QY 1769 ccaaggttcatgagacacacttaetgagatctgcccactcccagagagagagagagagagag 1828
 Db 1643 CCAAGGCTCCATGGACCGACGTGAGAGACTTTCCAAACCCCGGAGGAGGAACCTTCACA 1702

QY	1070	cattgcacccagggtagctgactgaatgagcaatcgtaccacctatcgtaaacctttagc	1129
DG	1167	CGCTGCAGCCCGCCTGGCTGGCGCATGGCAGGTCCGCTACCCCAATCGTACGCCCAGC	1226
QY	1130	caggcctgagtgagggecttgctgagtccaagactcttcctctttcccacaacaaact	1189
DB	1227	CAGCGCTGGCGTGFGGGCCCTGGCTGGCTCAAGACTCTCTTCCTCTTTCCCACCAGACC	1286
QY	1190	ggcttccccaatlaagcacagcgccttcacagctctactgctccgagactcgccagcct	1249
DB	1287	GGCTTCCCCAACAGTAGTACACCGCTCTCACGCTCTACTGCTTCGAGACTCTGCCAGGCC	1346
QY	1250	tctgcatatccttgagccttccacccagcctccacccagcctctgatgagactagaagct	1309
DB	1347	TCCACCACCCCTGAGCGCTCTGCA-----CCAGCCTCTGACGGCTCGAGGCC	1393
QY	1310	atcqtacaaatcacaaqacacctcgaggaactcgagctgcctcaggaagccacagaagtl	1369
DB	1394	ATTGTCACAGTGCACAGAGACCTTAGAGAGCTCCACGTCGTCGGCGAAGCTGTGGAGAC	1453
QY	1370	qaatcccgctggggccattactccatcccactcatggagagcggagagtgtagagctcc	1429
DB	1454	GAGTCCCGGGGAGCATCTACTCCGTCCTCCCATTTGTGGAGATGGGGAGGTGCAAGGTCC	1513
QY	1430	actcca	1435
DB	1514	CTCCA	1519

RESULT	14
I44675	
LOCUS	I44675 1519 bp DNA PAT
DEFINITION	Sequence 2 from patent US 563570.
ACCESSION	I44675
VERSION	I44675.1 GI:2469388
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1519)
TITLE	Hockfield,S. and Jaworski,D.M. DNA encoding BHAB, a brain hyaluronan-binding protein, and recombinant expression systems for production of BHAB polypeptides
JOURNAL	Patent: US 563570-A 2 03-JUN-1997;
FEATURES	Location/qualifiers 1..1519
source	/organism="unknown"
BASE COUNT	256 a 509 c 504 g 250 t
ORIGIN	

RESULT 15

